

# SCIENTIFIC REPORTS

OPEN

## Five genomic regions have a major impact on fat composition in Iberian pigs

R. N. Pena<sup>1</sup>, J. L. Noguera<sup>2</sup>, M. J. García-Santana<sup>3</sup>, E. González<sup>4</sup>, J. F. Tejeda<sup>4</sup>,  
R. Ros-Freixedes<sup>1,5</sup> & N. Ibáñez-Escriche<sup>6</sup>

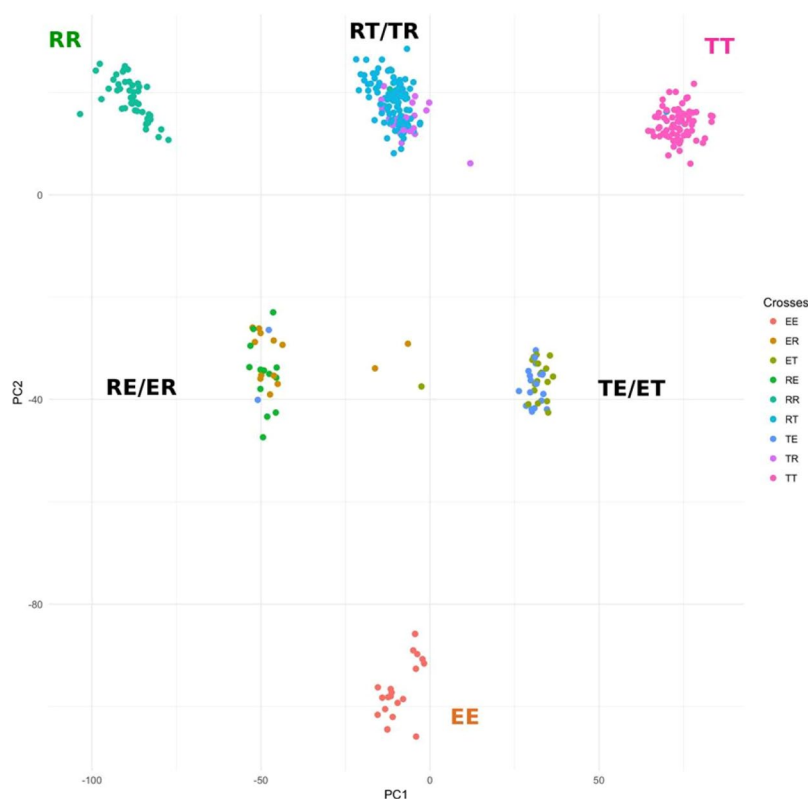
The adipogenic nature of the Iberian pig defines many quality attributes of its fresh meat and dry-cured products. The distinct varieties of Iberian pig exhibit great variability in the genetic parameters for fat deposition and composition in muscle. The aim of this work is to identify common and distinct genomic regions related to fatty acid composition in Retinto, Torbiscal, and Entrepelado Iberian varieties and their reciprocal crosses through a diallelic experiment. In this study, we performed GWAS using a high density SNP array on 382 pigs with the multimarker regression Bayes B method implemented in GenSel. A number of genomic regions showed strong associations with the percentage of saturated and unsaturated fatty acid in intramuscular fat. In particular, five regions with Bayes Factor >100 (SSC2 and SSC7) or >50 (SSC2 and SSC12) explained an important fraction of the genetic variance for miristic, palmitoleic, monounsaturated (>14%), oleic (>10%) and polyunsaturated (>5%) fatty acids. Six genes (*RXRB*, *PSMB8*, *CHGA*, *ACACA*, *PLIN4*, *PLIN5*) located in these regions have been investigated in relation to intramuscular composition variability in Iberian pigs, with two SNPs at the *RXRB* gene giving the most consistent results on oleic and monounsaturated fatty acid content.

The Iberian pig is a traditional breed native to the west and southwest of the Iberian Peninsula characterised by precocious development of fat depots and high deposition of intramuscular fat (IMF; ca. 8–12%)<sup>1</sup>, and small litter size<sup>2</sup>. The breed is mainly used for the production of high-quality dry-cured products, intended for niche, traditional or premium markets. In Iberian pigs, IMF is particularly rich in oleic acid (C18:1, n-9), a monounsaturated fatty acid that gives fat an oily aspect very appreciated by consumers. Pigs fed an acorn-based diet have around 55% of oleic acid content and relative low concentrations of linoleic and palmitic acids (around 8% and 20%, respectively)<sup>3</sup>. The breed is structured in 5 varieties (Ministry of Agriculture, Fishing and Food, Spain) which exhibit distinct growth, fattening and reproductive abilities. Pigs of the Torbiscal variety are characterised by a faster growing rate resulting in heavier carcasses and better noble cut yields than Entrepelado and Retinto<sup>4</sup>. On the other hand, the latter are more prolific than Torbiscal sows, their meat is darker and, although fattier, contains less saturated fat. These production differences indicate the segregation of variety-specific genetic variants, which, once known, could be combined to improve the productivity of the breed. Unlike most modern European breeds, the Iberian pigs do not show signs of admixture with Chinese breeds<sup>5</sup>, which emphasises the need to develop breeding strategies specific for the Iberian varieties. In a previous study<sup>6</sup> we confirmed the phenotypic differences between these three varieties. We also showed that fat deposition and the fatty acid profile of intramuscular (IMF) muscle and subcutaneous fat in Iberian pigs are traits of moderate or high heritability, indicating that selective breeding can be profitably used for the enhancement of such characteristics. The aim of this study is to use a high-density SNP genotyping platform to identify variety-specific variants associated with IMF content and composition through the analysis of a complete diallelic cross of three Iberian varieties.

<sup>1</sup>Departament de Ciència Animal, Universitat de Lleida-Agrotecnio Center, 25198, Lleida, Spain. <sup>2</sup>IRTA, Genètica i Millora Animal, 25198, Lleida, Spain. <sup>3</sup>INGA FOOD S.A, 06200, Almendralejo, Spain. <sup>4</sup>Tecnología de los alimentos, Universidad de Extremadura, 06006, Badajoz, Spain. <sup>5</sup>The Roslin Institute, Edinburgh University, Easter Bush, EH25 9RG, UK. <sup>6</sup>Institute for Animal Science and Technology, Universitat Politècnica de València, 46022, Valencia, Spain. Correspondence and requests for materials should be addressed to N.I.-E. (email: [noeibes@dca.upv.es](mailto:noeibes@dca.upv.es))

Male Variety. Boars used <sup>a</sup>	Female Variety. Sows used <sup>a</sup>		
	TT (81)	EE (24)	RR (39)
TT (18)	TT (77)	TE (21)	TR (26)
EE (12)	ET (17)	EE (21)	ER (13)
RR (14)	RT (106)	RE (13)	RR (48)

**Table 1.** Distribution of the pigs used in this study according to the parental and maternal Iberian variety used in each cross. Only the 342 offspring finally used in the GWAS analysis are indicated. <sup>a</sup>Varieties: TT = Torbiscal; EE = Entrepelado; RR = Retinto; RT = Retinto × Entrepelado; TR = Torbiscal × Retinto; RE = Retinto × Entrepelado, ER = Entrepelado × Retinto, TE = Torbiscal × Entrepelado; ET = Entrepelado × Torbiscal.



**Figure 1.** Population structure identified by principal component (PC) analysis. The three Iberian varieties (EE – Entrepelado; RR – Retinto; TT – Torbiscal) can be separated by the analysis, while the crossed pigs occupy intermediate positions.

## Results

**Phenotype differences between diallelic crosses.** A complete diallelic cross was set up by crossing pigs of three Iberian varieties, resulting in the generation of 9 different genotypes (Table 1). In this population we have previously shown that there is a genetic variety effect on the content of several IMF fatty acids<sup>6</sup>. In particular, the Torbiscal variety has more saturated fatty acids (SFA) and less monounsaturated fatty acids (MUFA) and MUFA/SFA ratio than Retinto pigs, while Entrepelado exhibits an intermediate phenotype (ref.<sup>6</sup> and Supplementary Table S1). Depending on the trait, the most relevant role is played by additive, heterosis and maternal effects or a combination of them. Here we have performed a genome-wide association analysis in an effort to identify the sources of these differences, focussing on IMF content and concentration of main saturated (SFA), monounsaturated (MUFA), and polyunsaturated (PUFA) fatty acids in *longissimus thoracis* muscle. After filtering the genotyping data, only 20,580 SNPs (less than one third) were retained for further analysis. A large fraction of the discarded SNPs did not segregate in Iberian pigs, which emphasises the need to develop breed-specific markers.

**Population structure.** Population stratification was assessed by a principal component analysis using the genotypic information of the 20,580 SNPs. The genetic diversity of the three pure varieties could be separated based on the PC results while the reciprocal crosses occupied intermediate positions between pure varieties following an expected pattern (Fig. 1). This information was used in the GWAS analysis to account for population stratification and control false positive results representing genetic differences from ancestry rather than genes associated with a trait<sup>7</sup>.

SSC <sup>a</sup>	Region <sup>b</sup>	n. SNP	Marker with highest BF	Trait <sup>c</sup>	BF	%GV	Candidate genes <sup>d</sup>
1	173.2–173.8	4	ALGA0007029	C14:0	11.53	4.13	
2	20.0–20.9	1	ALGA0109169	PUFA	81.97	4.95	<i>HSD17B12</i>
	75.0–75.6	2	MARC0048160	C16:1	624.06	14.44	<i>PLIN3</i> , <b><i>PLIN5</i></b> , <i>PLIN4</i> , <i>PIAS4</i> , <i>PIP5K1C</i> , <i>S1PR4</i> , <i>GNA15</i> , <i>GADD45B</i> , <i>CSNK1G2</i> , <i>ATP8B3</i> , <i>NDUFS7</i> , <i>STK11</i> , <i>ABCA7</i>
	90.4–97.9	4	DRGA0003183	C16:1	28.36	2.21	<i>SERINC5</i> , <i>ACOT12</i> , <i>MEF2C</i>
				C18:1	10.97	1.34	
				MUFA	13.45	2.25	
	100.1–106	4	ALGA0014789	MUFA/SFA	24.32	3.30	<i>NR2F1</i> , <i>ARSK</i>
				C16:1	52.33	4.07	
				C18:1	10.7	1.08	
				MUFA	12.63	1.33	
	115.1–115.7	3	MARC0058544	C14:0	16.54	1.35	<i>STARD4</i>
4	123.0–127.9	2	MARC0014258	MUFA/SFA	31.00	1.94	<i>AGL14</i> , <i>ABCD3</i> , <i>ABCA4</i> , <i>GFI1</i> , <i>SLC44A3</i>
6	3.0–4.0	2	ALGA0103513	C16:1	23.91	1	<i>FOXO2</i> , <i>COX4I1</i> , <i>MBTPS1</i>
7	25.0–25.9	2	H3GA0020505	C18:1	20.12	2.10	<i>APOM</i> , <i>ABHD16A</i> , <i>CYP21A2</i> , <i>SLC44A4</i> , <i>PPT2</i> , <b><i>PSMB8</i></b> , <i>HSD17B8</i> , <b><i>RXRβ</i></b> , <i>HMGCLL1</i>
	103.3–105.6	3	MARC0033066	C16:1	11.80	1.81	<i>TSHR</i> , <i>SEL1L</i>
	113.2–115.2	2	ASGA0036470	C18:1	117.66	10.40	<i>TTC7B</i> , <i>LGMN</i> , <b><i>CHGA</i></b> , <i>COX8C</i> , <i>SERPINA12</i> , <i>SERPINA6</i>
				MUFA	174.95	15.48	
				MUFA/SFA	117.36	6.13	
8	13.7–17.4	2	DRGA0008365	PUFA	37.36	3.26	<i>GBA3</i> , <i>PPARGC1A</i>
12	36.9–38.9	9	ASGA0054436	C14:0	77.66	17.13	<i>MED13</i> , <i>PIGW</i> , <b><i>ACACA</i></b>
14	137.5	1	DRGA001476	IMF	16.52	1.42	<i>EBF3</i>

**Table 2.** Genomic regions with strong associations with fatty acid content and compositional traits in *longissimus thoracis* muscle (Bayes factor (BF) > 10) and accounting for >1% genetic variance (%GV) in at least one trait. <sup>a</sup>SSC – *Sus scrofa* chromosome. <sup>b</sup>Region (in Mb) refer to pig genome assembly 11.1 coordinates. <sup>c</sup>Traits: Fatty acids, in % of total fatty acids; SFA – saturated fatty acid content; MUFA – monounsaturated fatty acid content; MUFA – monounsaturated fatty acid content; MUFA/SFA – MUFA to SFA ratio. <sup>d</sup>Positional candidate genes in bold have been investigated in this study. Gene full names in Supplementary Table S3.

**GWAS results.** We observed only one relevant association with IMF content at genomic level, located at pig chromosome (SSC) 14, 137.5 Mb, with a strong Bayes Factor (BF) of 16.52. This association explained a very minor fraction of the genetic variance of the trait (1.42%; Table 2). In addition, some suggestive regions were detected in SSC3 (44 Mb) and 9 (45 Mb) in relation to IMF content (Supplementary Fig. S1A). In contrast, 76 strong trait-SNP associations (Bayes Factor (BF) > 10) were found with at least one of the fat compositional traits. These involved 65 distinct SNPs that were associated with at least one of the fat compositional traits. In total, 32 distinct not overlapped regions (windows) distributed across 14 chromosomes were identified with BF > 10 (Supplementary Table S2), 15 of which explained at least 1% of the genetic variance for at least one trait (Table 2). Among them, there were two very strong association (BF > 100) involving four SNPs in SSC2 and SSC7. Additionally, two strong associations (BF > 50) were detected in SSC2 related to palmitoleic and PUFA content, plus nine markers associated with miristic acid content in SSC12. These five regions with the highest BF values also explained a large fraction of the genetic variance for miristic acid (SSC12), palmitoleic acid (SSC2, two regions), PUFA content (SSC2) and oleic acid and total MUFA content (SSC7), as detailed below.

In chromosome SSC2 we detected seven genomic regions strongly associated with fatty acid composition in the *longissimus thoracis* muscle (Supplementary Table S2). Three positional candidate genes, perilipin 3 (*PLIN3*), 4 (*PLIN4*) and 5 (*PLIN5*), map to the 75 Mb region, which showed the strongest association with palmitoleic acid content (BF = 624.1) and explained 14.44% of the genetic variance for this trait (Table 2). Two other regions in SSC2, at 90.4–97.9 Mb and 100.1–106 Mb (comprising 4 SNPs each), showed strong association with palmitoleic, oleic and MUFA content. The latter, in addition, was also associated to MUFA/SFA ratio. Given the relative close distance between these two regions, it cannot be ruled out that they might represent the same causal effect. Important genes in this region are *ACOT12* (89.6 Mb), a cytosolic thioesterase involved in the hydrolysis of medium- to long-chain fatty acyl-CoAs (C12–C18) to their free fatty acids, and *NR2F1* (100.4 Mb), a sterol-binding transcription factor (also known as COUP-TF1) involved in regulating the synthesis and transport of triglycerides in enterocytes<sup>8</sup>. Also of relevance in this chromosome is the association of the 20.0–20.9 Mb region with total PUFA content (BF = 81.97), which explained 4.95% of the genetic variance for this trait. A promising candidate gene in this region is *HSD17B12*. This gene encodes a 17 beta-hydroxysteroid dehydrogenase involved in estradiol synthesis in the ovary but also in fatty acid elongation in other tissues. The enzyme catalyses the second of the four reactions of the long-chain fatty acids elongation cycle towards very long-chain fatty acids, which are substrate of polyunsaturation events for instance by the fatty acid desaturases FADS2 and FADS3.

Animals	RXRB rs80789331	Difference in			
	genotypes compared	IMF content, %	SE	95% CI	p-value
Whole purebred population	CC-AA	1.61	0.53	0.36, 2.86	<b>0.008</b>
	CA-AA	1.01	0.43	0.00, 2.02	<b>0.050</b>
Torbiscal variety	CC-AA	2.15	0.74	0.39, 3.91	<b>0.012</b>
	CA-AA	1.23	0.56	−0.11, 2.58	0.079

**Table 3.** Effect of the *RXRB* exon 9 SNP rs80789331 on the intramuscular fat (IMF) content in *longissimus thoracis* muscle of Iberian pigs. SE – standard error; CI – confidence interval.

Two regions in SSC7, at positions 25.0–25.9 and 113.2–115.2 Mb, also displayed strong associations with more than one compositional trait (oleic acid and MUFA content and the MUFA/SFA ratio; Supplementary Table S2). Two SNPs at 113.2–115.2 Mb explained 10.40% and 15.48% of the genetic variance of oleic acid and MUFA content in intramuscular fat, respectively. Two interesting candidate genes are located at the 25 Mb region. The retinol-X-receptor beta (*RXRB*) gene is a strong functional candidate gene as it drives the lipogenic activities of retinoic acid, the bioactive vitamin A<sup>9</sup>. Another interesting gene is *PSMB8* which encodes a protease required for the differentiation of preadipocytes into adipocytes<sup>10</sup> and which disruption causes lipodystrophy<sup>11</sup>. The catestatin hormone, a product of the chromogranin A (*CHGA*) gene, which maps to SSC7, 114.3 Mb, is involved in the mobilisation of fatty acids from the adipose tissue. This gene encodes also pancreastatin, a hormone that regulates the insulin:glucagon ratio released from pancreas<sup>12</sup>. Therefore, *CHGA* represents a potential physiological candidate gene in this MUFA-associated region.

Finally, among the results it is also noteworthy a group of 9 close markers in the 36.9–38.9 Mb region of SSC12, all of them associated with miristic acid content. This association accounted for 17.13% of the genetic variance for miristic acid content, which represents the greatest effect detected in this study for a trait. This region co-localises with the acetyl-coA carboxylase (*ACACA*) gene (38.5–38.9 Mb), a gene that has been studied before in relation to muscle fatty acid composition in Iberian and Duroc pigs<sup>13–15</sup>.

**Functional evaluation of GWAS regions co-localising genes.** Genes mapping to 2-Mb windows around the 65 markers displaying strong associations (BF > 10) with one or more fatty acid compositional traits were catalogued using gene ontology terms and pathway participation in order to identify candidate genes whose function would be related (directly or indirectly) to fatty acid transport or metabolism. From the initial 974 genes, we identified 112 genes that participated in fatty acid biosynthesis, transport or catabolism, or that played a role in the differentiation and maturation of adipocytes (Supplementary Table S3). These constitute positional and functional candidate genes for the regions with BF > 10. Genes co-localising with regions explaining >1% of the genetic variance of at least one trait are indicated in Table 2. A subset of those genes were selected for further analysis considering if their molecular function was related to the trait affected in each region

**Characterisation of potential candidate genes.** We intended to sequence the proximal promoter and full cDNA of a group of 6 selected candidate genes in 4 Entrepelado, 4 Retinto and 4 Torbiscal pigs in order to identify polymorphisms segregating in these Iberian varieties. This list included *PLIN4* and *PLIN5* as candidate genes for the SSC2 75.0–75.6 Mb region explaining 14.44% of the genetic variance for palmitoleic acid content; *PSMB8* and *RXRB*, for the SSC7 25–25.9 Mb region for oleic acid content; *CHGA* for the SSC7 113.2–115.2 Mb region that explained more than 10% of the genetic variance for oleic and total MUFA content; and *ACACA* for the SSC12 36.9–38.9 Mb region for miristic acid content. The sequencing targets included exonic sequence and regulatory regions (promoters or putative enhancers). Unfortunately, we were not able to amplify the promoter or the exonic sequences of the *PLIN4* gene, which was hampered by the repetitive nature of this gene sequence. From the rest of the genes, a subset of the polymorphisms identified by Sanger sequencing on *PLIN5*, *RXRB*, *PSMB8*, *CHGA* and *ACACA* were selected to be genotyped by high resolution melting (HRM) PCR in 208 purebred pigs (28 Entrepelado, 72 Retinto and 108 Torbiscal) which included all purebred pigs from the GWAS experiment (n = 160) plus 48 additional pigs.

The most relevant result was with the *RXRB* polymorphisms. The two promoter polymorphisms (rs327226101 and rs331605500) were completely linked and in ample linkage disequilibrium (LD) with the distal SNP on exon 9 (LD = 0.94). The SNP on exon 9 (rs80789331) was associated with IMF content, with AA pigs having lower IMF levels than both CA and CC pigs (Table 3). This was an unexpected result as no signal for IMF was detected in this chromosome in the GWAS. The effect was particularly evident in the Torbiscal pigs (Table 3) where CC and AA pigs differed in 2.15% IMF content. This effect could not be captured in the other two varieties, which could be due to the lower number of pigs or it could indicate a variety-specific effect, due to a linked mutation in Torbiscal. However, the most significant associations for this gene were with compositional traits such as palmitic and oleic fatty acids, and also SFA, MUFA and PUFA content (Tables 4 and S4). The alleles that increased oleic acid content, MUFA and PUFA had a negative effect on SFA content. When the effects were studied in the individual Iberian varieties, the same effects were reproduced in Retinto animals but not in the others, suggesting a linked causal mutation specific of this variety (Tables 4 and S5). The magnitude of the effects was relevant between alternative homozygotes. For instance, for the *RXRB* rs327226101 promoter SNP, there was a 2% difference in oleic acid content between Retinto GG and TT pigs (Table 4) while differences in SFA and MUFA content were close to 4% and 3%, respectively.

Animals	Trait <sup>a</sup>	Genotypes compared	Difference in trait content	SE	95% CI	p-value
Whole purebred population	C16:0	TT-GG	0.80	0.27	0.15, 1.44	<b>0.011</b>
		TT-GT	0.56	0.20	0.08, 1.04	<b>0.018</b>
	C18:0	TT-GG	0.67	0.28	0.01, 1.32	<b>0.045</b>
		TT-GT	−0.85	0.40	−1.78, 0.09	0.086
	C18:1	TT-GG	−0.84	0.29	−1.54, −0.14	<b>0.013</b>
		TT-GT	1.47	0.49	0.32, 2.63	<b>0.008</b>
	SFA	TT-GG	0.97	0.36	0.11, 1.83	<b>0.023</b>
		TT-GT	−1.12	0.47	−2.22, −0.02	<b>0.046</b>
	MUFA	TT-GG	−0.94	0.35	−1.77, −0.12	<b>0.020</b>
		TT-GT	−0.37	0.15	−0.72, −0.03	<b>0.028</b>
	PUFA	TT-GG	−0.35	0.13	−0.67, −0.03	<b>0.026</b>
		TT-GT	−0.09	0.03	−0.16, −0.02	<b>0.013</b>
	MUFA/SFA	TT-GG	−0.06	0.02	−0.12, −0.01	<b>0.014</b>
		TT-GT	1.99	0.53	0.71, 3.28	<b>0.001</b>
Retinto variety	C16:0	TT-GG	1.51	0.43	0.48, 2.54	<b>0.002</b>
		TT-GT	−0.80	0.31	−1.53, −0.06	<b>0.032</b>
	C16:1	TT-GG	1.93	0.55	0.62, 3.25	<b>0.002</b>
		TT-GT	0.95	0.44	−0.11, 2.00	0.086
	C18:0	TT-GG	−2.28	0.78	−4.14, −0.41	<b>0.013</b>
		TT-GT	−1.98	0.62	−3.47, −0.48	<b>0.007</b>
	C18:1	TT-GG	3.90	1.02	1.44, 6.36	<b>0.001</b>
		TT-GT	2.52	0.82	0.55, 4.49	<b>0.009</b>
	SFA	TT-GG	−3.18	0.98	−5.53, −0.83	<b>0.005</b>
		TT-GT	−2.40	0.78	−4.28, −0.53	<b>0.009</b>
	MUFA	TT-GG	−0.75	0.21	−1.25, −0.25	<b>0.002</b>
		TT-GT	−0.62	0.19	−1.09, −0.16	<b>0.006</b>
	PUFA	TT-GG	−0.23	0.07	−0.39, −0.07	<b>0.002</b>
		TT-GT	−0.16	0.05	−0.29, −0.04	<b>0.008</b>

**Table 4.** Effect of the *RXRB* promoter SNP rs327226101 on the intramuscular fat composition in *longissimus thoracis* muscle of Iberian pigs analysed together or by variety. <sup>a</sup>All traits are expressed as % of total fatty acid content. SFA - saturated fatty acids; MUFA - monounsaturated fatty acids; PUFA - polyunsaturated fatty acids; SE - Standard error.

Results in other mutations were less consistent. For instance, the *PSMB8* rs80801731 promoter mutation was significantly associated with oleic acid content (Supplementary Table S4), a trait that was selected for using *PSMB8* as a candidate gene. However, this association was not detected in any of the three varieties when analysed separately, which might indicate that this is a spurious result. Of the two mutations of the *CHGA* gene, only the rs322866746 showed significant differences between AG and GG pigs on stearic acid, oleic acid, SFA and MUFA content (Supplementary Tables S4 and S5) when analysed in the whole purebred population and also in the Entrepelado variety. The magnitude of the effects in Entrepelado were lower than those described above for the *RXRB* SNPs. On the other hand, the effects of the *CHGA* rs318914549 SNP on IMF composition (oleic acid, SFA and MUFA) were only detected in Retinto animals, with CC and CT pigs having ~2.5% and ~3% more oleic acid and MUFA than TT pigs, respectively. The opposite effect was seen in SFA content, probably as an indirect effect on the percentage content. Finally, three SNPs in *ACACA* were selected as candidate genes for miristic acid content. However, none of them was associated with this trait. Instead, the 12:38727103 G > A mutation in a putative *ACACA* enhancer on intron 24 was associated with palmitic, oleic, SFA and MUFA content both in the whole population and in the Retinto variety (Supplementary Tables S4 and S5). In addition, no significant associations were found between a missense mutation in the exon 3 of the *PLIN5* gene (Supplementary Table S6) and any of the compositional traits.

## Discussion

We have performed a GWAS of three distinct Iberian pig varieties and their reciprocal crosses to identify relevant genomic regions associated with IMF content and composition in these particular varieties. This is the first GWAS performed in pure variety Iberian animals, which makes results very valuable for breeders working with this breed. Despite losing nearly 2/3 of the potential genotypes in the high-density panel due to lack of segregation of these SNPs, we were able to detect 32 genomic regions with strong associations with several fatty acid compositional traits. Out of these, two regions stand out by their ability to capture a very significant fraction of the genetic variance of monounsaturated fatty acids. These are the 75–75.6 Mb region in SSC2, which explains 14.4% of the genetic variance for palmitoleic acid content, the 113.2–115.2 Mb region in SSC7 (10% and 15% of the genetic variance of oleic acid and MUFA content, respectively). Two additional regions in SSC2 captured about



4–5% of the genetic variance for PUFA (20.0–20.9 Mb) and palmitoleic acid (100.1–106 Mb), while the SSC12 36.9–38.9 Mb region explained 17% of the genetic variance of miristic acid content.

Oleic and palmitoleic acid are the two main MUFA in pork. In Iberian meat, usual percentages are in the range of 45–55% and 4–5%, respectively. Monounsaturated fat represents a healthier alternative to the consumption of SFA, without the rancidity odour risk of PUFA<sup>16</sup>. In Iberian cured products the content of oleic acid contributes to the taste and visual aspect sought after by consumers which are willing to pay more for a high-quality product. This puts into context the relevance of identifying genomic regions which explain a large fraction of the genetic variance for the main MUFA in pork.

We have studied 6 genes located in some of these regions. All of them have a molecular function related to fatty acid metabolism (*RXRB*, *ACACA*), storage (*PLIN4*, *PLIN5*, *CHGA*) or adipogenesis (*PSMB8*). After sequencing coding and regulatory regions, a subset of SNP markers was selected to be genotyped in a group of 208 purebred Iberians from the Entrepelado, Retinto and Torbiscal varieties. The most consistent result was observed for three partially linked SNPs in the *RXRB* gene (two in the promoter and one at the end of the gene, on exon 9). Considering the rs80789331 SNP on exon 9, pigs of the CC genotype had more SFA and less MUFA and oleic acid than either AA or CA pigs. *RXRB* is an ubiquitously expressed transcription factor from the nuclear receptor superfamily also known as NR2B2. By binding 9-*cis*-retinoid acid, it can integrate a number of metabolic signalling pathways related to energy metabolism by forming homodimers or, more frequently, heterodimers with other nuclear receptor members such as RAR $\alpha,\beta,\gamma$ , RXR $\alpha,\gamma$  or PPAR $\alpha,\beta,\delta/\gamma$ <sup>17</sup>. In liver, activation of *RXRB* increases the expression of stearoyl-coA desaturase (*SCD*) and *CD36* fatty acid translocase<sup>18</sup> and *in vitro* adipocyte differentiation<sup>17</sup>. *CD36* is a membrane transporter very abundant in skeletal muscle that participates in the uptake of long- and very long-chain fatty acids. The *SCD* enzyme, on the other hand, catalyses the limiting step in the synthesis of monounsaturated fatty acids (oleic and palmitoleic acids) from saturated substrates (stearic and palmitic acids). A mutation in the *SCD* proximal promoter has been associated to higher oleic acid content in Duroc pork without affecting the total fat content in muscle or backfat<sup>19</sup>. However, this mutation does not segregate in Iberian pigs (Estany *et al.*<sup>19</sup> and own data on our Iberian resource population). It is likely that the effect of *RXRB* polymorphisms on MUFA content of the Iberian pigs is driven by differential expression on the *SCD* gene and/or other genes related to fatty acid elongation, desaturation or oxidation. However, one must be cautious with assigning functional relevance only based on the position of a mutation. The most proximal rs327226101 mutation lays –35bp from the transcriptional start site of *RXRB*, and could therefore affect change the functionality of the core promoter of this gene. However, on the other strand this mutation lays in first exon/intron of the two transcriptional variants of the *SLC39A7*, which encodes for a Zn<sup>2+</sup> Golgi transporter. A simultaneous effect of the rs327226101 SNP on either or both genes cannot be ruled out. Future analysis of muscle transcriptomics will provide complementary information to the GWAS results.

The conjoint effect of the *RXRB* SNPs on SFA, MUFA and oleic acid content was also observed for the rs322866746 and rs318914549 mutations on exon 6 of the *CHGA* gene and for an intron 5 mutation in *ACACA*, and it is due to the high genetic correlation between these traits (positive between MUFA and oleic acid and negative between SFA and MUFA or oleic)<sup>20</sup>. Chromogranin A is a secretory proprotein expressed from the *CHGA* gene that is cleaved in several active peptidic hormones including pancreastatin, vasostatin, and catestatin<sup>12</sup>. The latter is a pleiotropic hormone that has effects on promoting angiogenesis, lowering blood pressure, immunomodulation, and also has a negative effect on fat deposition by increasing fatty acid oxidation<sup>12</sup>. Although this is a very interesting gene, the two missense mutations we found overlapped with none of the biologically active peptides, ruling out the causality of these mutations on the IMF composition. Also, a promising mutation in exon 3 of the *PLIN5* gene which promoted a non-conservative change at the protein sequence (Supplementary Table S6), was not associated with intramuscular fatty acid composition. Still, the relevance of this gene regarding monounsaturated fatty acid content has been highlighted recently, in relation to the activity of the hormone-sensitive lipase LIPE<sup>21</sup>. Therefore, this is a gene to be pursued in future experiments.

The acetyl-coA carboxylase alpha (*ACACA*) gene was selected as a physiological and positional candidate gene for the region on SSC12 (38 cM) for miristic acid content. The *ACACA* enzyme is a limiting factor in the synthesis of long chain fatty acids and plays an important role in the conversion of acetyl-CoA to malonyl-CoA, acting as a precursor of miristic and palmitate acids<sup>13</sup>. In addition, its activity slows the importation of fatty acids to the mitochondria for  $\beta$ -oxidation. Malonyl-CoA, as a product of the enzyme *ACACA*, is the intermediate substrate of other enzymes that will synthesize SFA and MUFA<sup>15</sup>. In pigs, *ACACA* is a 287 Kb-long gene structured in 56 exons, which is transcribed into five variants that differ in the use of certain exons and with a complex regulation from several promoters<sup>15</sup>. Several polymorphisms have been described in both coding and regulatory regions of this gene, some of which have been previously studied in Iberian pigs or in experimental crosses with Iberian pigs<sup>15</sup>. A synonymous polymorphism in exon 49 (c.5634 T > C; rs340781986), likely linked to a yet unidentified causal mutation, has been associated with the percentage of several fatty acids in crossed Iberian pigs<sup>14,15</sup> and with lean and intramuscular fat content in Duroc<sup>13</sup>. We tested this mutation in our diallelic population along with two mutation in a putative enhancer in intron 24 of the gene (based on human regulatory data). In our Iberian population the synonymous rs340781986 SNP in exon 49 showed no effect on the fatty acid compositional data, confirming lack of causality. In contrast, one of the enhancer SNPs (12:38727103 G > A) was significantly associated with several SFA and MUFA fatty acids, but not with miristic content. Given the complexity of the gene, it might be challenging to demonstrate causality related to *ACACA* activity, and rule out other close polymorphisms in linkage disequilibrium with the tested mutations.

Taken together, the results presented here show that five genomic regions at SSC2, 7 and 12 regulate a relevant percentage of the genetic variance for fatty acid composition in Iberian pigs. It is relevant that most results presented here were specific of the Retinto variety (except the *CHGA* rs32866746 SNP, which was specific of Entrepelado). Retinto pigs represented 72 of the 208 pigs in the resource population and are characterised by an enhanced deposition of both IMF and backfat, which is higher in MUFA and oleic acid than in the other two

varieties. These results most probably indicate the existence of Retinto/Entrepelado-specific causal mutations linked to the SNPs analysed here. This would agree with the segregation of genetic variety effects on IMF content and composition described in these Iberian varieties<sup>6</sup>. In the path to develop genetic markers useful for Iberian breeders, the results presented here support further study in the variety-specific genetic variability by means of targeted genome sequencing and global muscle transcriptomics.

## Materials and Methods

**Ethics approval.** All animal procedures followed Spanish national guidelines and were approved by the Ethical Committee of Institut de Recerca i Tecnologia Agroalimentàries (IRTA). The animal experiments were carried out in accordance with the Guidelines for Animal Experiments.

**Experimental design.** The animal material is described in detail by Ibáñez-Escriche *et al.*<sup>6</sup>. Briefly, a complete diallelic cross was generated at the Inga Food S.A. facilities involving three varieties of Iberian pigs (Torbiscal, Retinto, and Entrepelado) and their reciprocal crosses (Retinto × Entrepelado, Entrepelado × Retinto, Torbiscal × Retinto, Retinto × Torbiscal, Torbiscal × Entrepelado, and Entrepelado × Torbiscal). The 470 pigs generated were reared in standard intensive commercial conditions and had *ad libitum* access to typical commercial feed used in Iberian pig production<sup>6</sup>. The animals were slaughtered in a commercial slaughterhouse at an average of 340 days of age (SD 14.34) and 160.03 kg (SD 11.60) of live weight. At slaughter, a section of the *longissimus thoracis* muscle was collected for fatty acid content and composition determination. Samples of liver were also collected, snap-frozen in liquid nitrogen and stored at −80 °C for subsequent nucleic acid isolation.

**Fatty acid content and composition.** For each animal, a sample of 200 g of *longissimus thoracis* muscle was minced and homogenized using a horizontal cutter mixer. Total IMF and its fatty acid profile were quantified according to the method described by O'Fallon<sup>22</sup>. Descriptive statistics and abbreviations for the traits analysed are included in Supplementary Table S1.

**DNA and RNA isolation.** DNA was isolated from liver samples using standard protocols as described by Ros-Freixedes *et al.*<sup>23</sup>. RNA was isolated from liver and muscle samples from a subset of the pigs in the experiment by the acid phenol method<sup>24</sup> using TRI-reagent (Sigma-Aldrich) and following the manufacturer's instructions.

**Principal Components Analysis.** Principal Components Analysis (PCA) with the SNP genotypes of the individuals passing the genotyping quality control was analysed with *prcomp()* base function implemented in R and used to infer population structure in genetic data and to check for outliers.

**Genome-wide association analysis for meat quality traits.** Genotyping was performed with the Porcine SNP60 BeadChip (Illumina), which contains 62,163 single nucleotide polymorphisms (SNPs). Quality control of data was carried out with the GenomeStudio software (Illumina), removing SNPs which either (i) mapped to the X chromosome, (ii) had a missing genotypes rate >5%, (iii) did not conform Hardy-Weinberg expectations (threshold set at a P-value ≤ 0.001), (iv) had a minor allele frequency below 0.05, (v) showed a GenCall score <0.15, (vi) had a call rate <95% or (vii) could not be mapped to the Sscrofa 11.1 version of the pig genome. After filtering the raw data, a GWAS was carried out with 20,580 SNPs from 342 individuals. Multi-SNP association analyses were performed with the Bayes B multiple marker regression (MMR) method<sup>25</sup> implemented in the GenSel software<sup>26</sup> under an additive genetic model. The kinship matrix was not included in the model because the makers fitted in the Bayesian MMR models capture the kinship structure<sup>27</sup>. We included the two first principal components to avoid possible patterns of population stratification. Nevertheless, Bayesian MMR models seems to be robust to population structure and to relationships<sup>28</sup>.

The statistical model assumed in this analysis was:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_{j=1}^k \mathbf{z}_j \alpha_j \delta_j + \mathbf{e},$$

where  $\mathbf{y}$  is the phenotype vector  $n \times 1$ ;  $\mathbf{X}$  is the incidence matrix of systematic effects  $n \times p$ ;  $\mathbf{b}$  is the systematic effect vector that includes the two first principal components (52% and 32% of the variance, respectively) as covariates, sex (2 levels) and batch (12 levels) effects and weight at slaughter as covariate for IMF trait, or the IMF content as covariate for fatty acid compositional data;  $\mathbf{z}_j$  is the genotype vector  $n \times 1$  for each SNP in the  $j$  locus ( $j = 1$  to  $k$ , where  $k$  is the SNP number);  $\alpha$  is the allelic substitution vector for each SNP in the  $j$  locus;  $\delta$  is a 0/1 random variable that indicates the absence or presence in the model (using  $\pi$  and  $1 - \pi$  as prior probabilities, respectively) of the  $j$  SNP for each iteration of the Markov-Monte Carlo chain; and  $\mathbf{e}$  is the residual vector assumed to follow a normal distribution. Due to the limited number of animals in this study, the proportion of SNPs without effect in the model ( $\delta = 0$ ) was established at  $\pi = 0.995$ . Variance components calculated in a previous study<sup>6</sup> were used as priors in the analyses. The statistical relevance of the associations between individual markers and traits was assessed by Bayes Factor (BF) for each locus<sup>29</sup>. Evidence of association was considered important for a BF > 5, strong when > 10 and decisive when > 100<sup>30</sup>. The proportion of variance explained conjointly by SNPs in not overlapped windows was estimated.

**Mapping SNPs to genes and pathways.** Positional candidate genes up to 2 Mb around the markers with BF > 10 for at least one compositional trait were retrieved from Ensembl (EMBL-EBI) by Biomart. Functional annotation of the genes was performed with several tools with the aim to catalogue and identify genes with a potential role in fatty acid synthesis, transformation, storing, oxidation or transport. Tools included

Enrichr<sup>31</sup>, which integrate several Gene Ontology and Pathway Analysis tools such as KEGG, Reactome and Wikipathways; Generic Gene Ontology Term Mapper (<http://go.princeton.edu/cgi-bin/GOTermMapper>); DAVID Bioinformatics Resources 6.8<sup>32</sup>; and the Go-Slim functional classification features of Panther (<http://pantherdb.org/>). More information can be found in Supplementary Table S3.

**Candidate gene analysis.** The transcribed region and approximately 2 kb of 5' flanking genomic sequence was amplified from 5 selected genes (*PLIN4*, *PLIN5*, *RXR*, *PSRMB8*, *CHGA*). As the *ACACA* coding and promoter regions have been well characterised in Iberian and Duroc pigs, we selected a 1.8-kb fragment in intron 24 of the gene which corresponds to the sequence of an internal human *ACACA* enhancer described in the enhancer-to-gene associations in GeneHancer (embedded in GeneCards). Sequences were retrieved from the Sscrofa 10.2 assembly in Ensembl and primer pairs were designed using Primer3Plus<sup>33</sup> (Supplementary Table S5). PCR conditions are described in detail Supplementary Table S6. Amplicons were purified with Exonuclease I and Shrimp alkaline phosphatase (ThermoFisher) according to the manufacturer's instructions prior to Sanger sequencing. Electropherograms were assembled and processed with GeneStudio Pro (2.2.0.0) (GeneStudio, Inc.). Genotyping of polymorphisms was performed by high-resolution melting protocols in QuantStudio QS3 real-time thermocycler (LifeTechnologies) with primers and conditions described in Supplementary Table S6.

**Association analysis.** In order to perform association analysis between IMF content or FA composition and the genotyped SNPs in the five candidate genes, the following standard linear model was used:

$$y_{ijklm} = \mu + batch_j + Sex_k + Variety_l + \beta * Weight_{ijklm} + Genotype_m + e_{ijklm}$$

where  $y_{ijklm}$  is the vector of phenotypic observations i.e. (IMF content; fatty acid composition) measured at the *longissimus thoracis* muscles of the  $i^{th}$  individual;  $\mu$  is the population mean of each trait;  $batch_j$  is a systematic effect of the  $j^{th}$  fattening batch, with 11 categories;  $Sex_k$  is the sex (castrated male/female) of the  $i^{th}$  individual;  $\beta$  is the regression coefficient on the covariate weight at slaughter (for IMF) or IMF content (for fatty acid composition) ( $Weight_{ijklm}$ );  $Variety_l$  is the effect of the variety, with 3 levels (Retinto, Torbiscal, Entrepelado);  $Genotype_m$  is the genotype for each SNP, with three levels; and  $e_{ijklm}$  is the residual effect of the  $i^{th}$  individual. Additionally, the same model was applied to each individual Iberian variety, removing the  $Variety_l$  effect from the model. Differences between genotype means were assessed with a Tukey test (or a two-tail t-test when only two genotypes were available). Means were considered different at p-value <0.05.

## Data Availability

The results from all data analysed during this study are included or displayed in this article (and its Supplementary Information files, such as the Manhattan plots). Associations for individual SNPs with  $BF > 10$  are tabulated (Supplementary Table S2) and regionally displayed (Supplementary Fig. S1). However, other individual SNP results may be made available from the corresponding author on request.

## References

- García Casco, J. M., Muñoz Muñoz, M., SilioLopez, L. & Rodríguez Valdovinos, C. Genotype by environment interaction for carcass traits and intramuscular fat content in heavy Iberian pigs fattened in two different free-range systems. *Span J Agric Res* **12**, 388–395, <https://doi.org/10.5424/sjar/2014122-4840> (2014).
- Lopez-Bote, C. J. Sustained utilization of the Iberian pig breed. *Meat Sci* **49**, S17–S27 (1998).
- Benito, J., Albarran, A. & García Casco, J. M. In Sustainable grassland productivity: Proceedings of the 21st General Meeting of the European Grassland Federation, Badajoz, Spain, 3–6 April, 2006 (eds Lloveras, J. et al.) 635–645 (Sociedad Española para el Estudio de los Pastos(SEEP), 2006).
- Benito, M. J., Rodríguez, M. M., Córdoba, M. G., Aranda, E. & Córdoba, J. J. Rapid differentiation of *Staphylococcus aureus* from staphylococcal species by arbitrarily primed-polymerase chain reaction. *Let Appl Microbiol* **31**, 368–373, <https://doi.org/10.1046/j.1472-765x.2000.00833.x> (2000).
- Groenen, M. A. A decade of pig genome sequencing: a window on pig domestication and evolution. *Genet Sel Evol* **48**, 23, <https://doi.org/10.1186/s12711-016-0204-2> (2016).
- Ibanez-Escriche, N., Magallon, E., Gonzalez, E., Tejeda, J. F. & Noguera, J. L. Genetic parameters and crossbreeding effects of fat deposition and fatty acid profiles in Iberian pig lines. *J Anim Sci* **94**, 28–37, <https://doi.org/10.2527/jas.2015-9433> (2016).
- Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909, <https://doi.org/10.1038/ng1847> (2006).
- Dai, K., Khatun, I. & Hussain, M. M. NR2F1 and IRE1beta suppress microsomal triglyceride transfer protein expression and lipoprotein assembly in undifferentiated intestinal epithelial cells. *Arterioscler Thromb Vasc Biol* **30**, 568–574, <https://doi.org/10.1161/ATVBAHA.109.198135> (2010).
- Wang, B. et al. Nutrigenomic regulation of adipose tissue development - role of retinoic acid: A review. *Meat Sci* **120**, 100–106, <https://doi.org/10.1016/j.meatsci.2016.04.003> (2016).
- Arimochi, H., Sasaki, Y., Kitamura, A. & Yasutomo, K. Differentiation of preadipocytes and mature adipocytes requires PSMB8. *Sci Rep* **6**, 26791, <https://doi.org/10.1038/srep26791> (2016).
- Kitamura, A. et al. A mutation in the immunoproteasome subunit PSMB8 causes autoinflammation and lipodystrophy in humans. *J Clin Invest* **121**, 4150–4160, <https://doi.org/10.1172/JCI58414> (2011).
- Bandyopadhyay, G. K. et al. Catestatin (chromogranin A(352–372)) and novel effects on mobilization of fat from adipose tissue through regulation of adrenergic and leptin signaling. *J Biol Chem* **287**, 23141–23151, <https://doi.org/10.1074/jbc.M111.335877> (2012).
- Gallardo, D. et al. Polymorphism of the pig acetyl-coenzyme A carboxylase alpha gene is associated with fatty acid composition in a Duroc commercial line. *Anim Genet* **40**, 410–417, <https://doi.org/10.1111/j.1365-2052.2009.01854.x> (2009).
- Muñoz, G. et al. QTL detection on porcine chromosome 12 for fatty-acid composition and association analyses of the fatty acid synthase, gastric inhibitory polypeptide and acetyl-coenzyme A carboxylase alpha genes. *Anim Genet* **38**, 639–646, <https://doi.org/10.1111/j.1365-2052.2007.01668.x> (2007).
- Munoz, M. et al. Disentangling Two QTL on Porcine Chromosome 12 for Backfat Fatty Acid Composition. *Anim Biotechnol* **24**, 168–186, <https://doi.org/10.1080/10495398.2012.763130> (2013).



16. Webb, E. C. & O'Neill, H. A. The animal fat paradox and meat quality. *Meat Sci* **80**, 28–36, <https://doi.org/10.1016/j.meatsci.2008.05.029> (2008).
17. Hamza, M. S. *et al.* De-novo identification of PPARgamma/RXR binding sites and direct targets during adipogenesis. *PLoS One* **4**, e4907, <https://doi.org/10.1371/journal.pone.0004907> (2009).
18. Singh Ahuja, H. *et al.* Differential effects of rexinoids and thiazolidinediones on metabolic gene expression in diabetic rodents. *Mol Pharmacol* **59**, 765–773 (2001).
19. Estany, J., Ros-Freixedes, R., Tor, M. & Pena, R. N. A functional variant in the stearoyl-CoA desaturase gene promoter enhances fatty acid desaturation in pork. *PLoS ONE* **9**, e86177, <https://doi.org/10.1371/journal.pone.0086177> (2014).
20. Ros-Freixedes, R., Reixach, J., Bosch, L., Tor, M. & Estany, J. Genetic correlations of intramuscular fat content and fatty acid composition among muscles and with subcutaneous fat in Duroc pigs. *J Anim Sci* **92**, 5417–5425, <https://doi.org/10.2527/jas.2014-8202> (2014).
21. Zappaterra, M. *et al.* Investigation of the Perilipin 5 gene expression and association study of its sequence polymorphism with meat and carcass quality traits in different pig breeds. *Animal* **12**, 1135–1143, <https://doi.org/10.1017/S1751731117002804> (2018).
22. O'Fallon, J. V., Busboom, J. R., Nelson, M. L. & Gaskins, C. T. A direct method for fatty acid methyl ester synthesis: application to wet meat tissues, oils, and feedstuffs. *J Anim Sci* **85**, 1511–1521, <https://doi.org/10.2527/jas.2006-491> (2007).
23. Ros-Freixedes, R. *et al.* Genome-Wide Association Study Singles Out SCD and LEPR as the Two Main Loci Influencing Intramuscular Fat Content and Fatty Acid Composition in Duroc Pigs. *PLoS One* **11**, e0152496, <https://doi.org/10.1371/journal.pone.0152496> (2016).
24. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**, 156–159, <https://doi.org/10.1006/abio.1987.9999> (1987).
25. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
26. Fernando, R. L. & Garrick, D. J. GenSel: User Manual for a Portfolio of Genomic Selection Related Analyses. (ed. Iowa State University Animal Breeding and Genetics, Ames.) (2009).
27. Habier, D., Fernando, R. L. & Dekkers, J. C. M. The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* **177**, 2389–2397 (2007).
28. Tossi, R., Fernando, R. L. & Dekkers, J. C. M. Genome-wide mapping of quantitative trait loci in admixed populations using mixed linear model and Bayesian multiple regression analysis. *Genetics Selection Evolution* **50**, 32 (2018).
29. Legarra, A. *et al.* A comparison of methods for whole-genome QTL mapping using dense markers in four livestock species. *Genet Sel Evol* **47**, 6, <https://doi.org/10.1186/s12711-015-0087-7> (2015).
30. Kass, R. E. & Raftery, A. E. Bayes Factors. *Journal of the American Statistical Association* **90**, 773–795 (1995).
31. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–97, <https://doi.org/10.1093/nar/gkw377> (2016).
32. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
33. Untergasser, A. *et al.* Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**, W71–74, <https://doi.org/10.1093/nar/gkm306> (2007).

## Acknowledgements

The authors would like to acknowledge the help of Inga Food S.A. and their support staff in setting up the experiment (E. Magallón, M. Ramos, L. Muñoz, and P. Diaz) and Dr L. Varona from University of Zaragoza for discussion. This study was funded by the INIA grant RTA 2012-0054-C02-01 and the Centre for the Development of Industrial Technology agreement IDI-20140447.

## Author Contributions

N.I.E., J.L.N. designed the experiment. J.L.N., M.J.G.S. generated the experimental cross. E.G., J.F.T. analysed the fatty acid profile. N.I.E., R.R.F. performed the analysis of the genomic data. R.N.P. investigated the candidate genes. All authors participated in the writing and discussion of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-38622-7>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019